

RAW SEQUENCE LISTING  
 PATENT APPLICATION: US/09/965,807  
 DATE: 11/19/2001  
 TIME: 14:24:06

Input Set : N:\Crif3\RULE60\09965807.txt  
 Output Set: N:\CRF3\11192001\I965807.raw

SEQUENCE LISTING

4 (1) GENERAL INFORMATION:  
 6 (i) APPLICANT: Matalon, Reuben  
 7 Kaul, Rajinder  
 8 Gao, Guang Ping  
 9 Balamurugan, Kuppareddi  
 10 Michals-Matalon, Kimberlee  
 12 (ii) TITLE OF INVENTION: Aspartoacylase Gene, Protein, and  
 13 Methods of Screening for Mutations Associated with  
 Canavan  
 14 Disease  
 16 (iii) NUMBER OF SEQUENCES: 27  
 18 (iv) CORRESPONDENCE ADDRESS:  
 19 (A) ADDRESSEE: Millen, White, Zelano & Branigan, P.C.  
 20 (B) STREET: 2200 Clarendon Boulevard, Suite 1400  
 21 (C) CITY: Arlington  
 22 (D) STATE: Virginia  
 23 (E) COUNTRY: U.S.A.  
 24 (F) ZIP: 22201  
 26 (v) COMPUTER READABLE FORM:  
 27 (A) MEDIUM TYPE: Floppy disk  
 28 (B) COMPUTER: IBM PC compatible  
 29 (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
 30 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25  
 32 (vi) CURRENT APPLICATION DATA:  
 C--> 33 (A) APPLICATION NUMBER: US/09/965,807  
 C--> 34 (B) FILING DATE: 01-Oct-2001  
 40 (C) CLASSIFICATION:  
 37 (vii) PRIOR APPLICATION DATA:  
 38 (A) APPLICATION NUMBER: US 08/128,020  
 39 (B) FILING DATE: 29-SEP-1993  
 42 (viii) ATTORNEY/AGENT INFORMATION:  
 43 (A) NAME: Hamlet-King, Diana  
 44 (B) REGISTRATION NUMBER: 33,302  
 45 (C) REFERENCE/DOCKET NUMBER: Shutt 1  
 47 (ix) TELECOMMUNICATION INFORMATION:  
 48 (A) TELEPHONE: 703-243-6333  
 49 (B) TELEFAX: 703-243-6410  
 50 (C) TELEX: 64191  
 53 (2) INFORMATION FOR SEQ ID NO: 1:  
 55 (i) SEQUENCE CHARACTERISTICS:  
 56 (A) LENGTH: 1435 base pairs  
 57 (B) TYPE: nucleic acid  
 58 (C) STRANDEDNESS: double  
 59 (D) TOPOLOGY: linear  
 62 (ix) FEATURE:  
 63 (A) NAME/KEY: CDS  
 64 (B) LOCATION: 159..1097

ENTERED

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/965,807

DATE: 11/19/2001

TIME: 14:24:06

Input Set : N:\Crif3\RULE60\09965807.txt

Output Set: N:\CRF3\11192001\I965807.raw

70 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

72	TTGTAACAGA	AAATTAAAT	ATACTCCACT	CAAGGGAATT	CTGTACTTTG	CCCTTTTGGT	60
74	AAAGTCTCAT	TTACATTTCT	AAACCTTTCT	TAAGAAAATC	GAATTCCTT	TGATCTCTCT	120
76	TCTGAATTGC	AGAAATCAGA	TAAAACTAC	TTGGTGAA	ATG ACT	TCT TGT CAC	173
77					Met Thr Ser Cys His		
78				1		5	
80	ATT GCT GAA	GAA CAT ATA	CAA AAG GTT	GCT ATC TTT	GGA GGA ACC	CAT	221
81	Ile Ala Glu	Glu His Ile	Gln Lys Val	Ala Ile Phe	Gly Gly Thr	His	
82		10		15		20	
84	GGG AAT GAG	CTA ACC GGA	GTA TTT CTG	GTT AAG CAT	TGG CTA GAG	AAT	269
85	Gly Asn Glu	Leu Thr Gly	Val Phe Leu	Val Lys His	Trp Leu Glu	Asn	
86		25		30		35	
88	GGC GCT GAG	ATT CAG AGA	ACA GGG CTG	GAG GTA AAA	CCA TTT ATT	ACT	317
89	Gly Ala Glu	Ile Gln Arg	Thr Gly Leu	Glu Val Lys	Pro Phe Ile	Thr	
90		40		45		50	
92	AAC CCC AGA	GCA GTG AAG	AAG TGT ACC	AGA TAT ATT	GAC TGT GAC	CTG	365
93	Asn Pro Arg	Ala Val Lys	Lys Cys Thr	Arg Tyr Ile	Asp Cys Asp	Leu	
94		55		60		65	
96	AAT CGC ATT	TTT GAC CTT	GAA AAT CTT	GGC AAA AAA	ATG TCA GAA	GAT	413
97	Asn Arg Ile	Phe Asp Leu	Glu Asn Leu	Gly Lys Lys	Met Ser Glu	Asp	
98	70		75		80		85
100	TTG CCA TAT	GAA GTG AGA	AGG GCT CAA	GAA ATA AAT	CAT TTA TTT	GGT	461
101	Leu Pro Tyr	Glu Val Arg	Arg Ala Gln	Glu Ile Asn	His Leu Phe	Gly	
102		90		95		100	
104	CCA AAA GAC	AGT GAA GAT	TCC TAT GAC	ATT ATT TTT	GAC CTT CAC	AAC	509
105	Pro Lys Asp	Ser Glu Asp	Ser Tyr Asp	Ile Ile Phe	Asp Leu His	Asn	
106		105		110		115	
108	ACC ACC TCT	AAC ATG GGG	TGC ACT CTT	ATT CTT GAG	GAT TCC AGG	AAT	557
109	Thr Thr Ser	Asn Met Gly	Cys Thr Leu	Ile Leu Glu	Asp Ser Arg	Asn	
110		120		125		130	
112	AAC TTT TTA	ATT CAG ATG	TTT CAT TAC	ATT AAG ACT	TCT CTG GCT	CCA	605
113	Asn Phe Leu	Ile Gln Met	Phe His Tyr	Ile Lys Thr	Ser Leu Ala	Pro	
114		135		140		145	
116	CTA CCC TGC	TAC GTT TAT	CTG ATT GAG	CAT CCT TCC	CTC AAA TAT	GCG	653
117	Leu Pro Cys	Tyr Val Tyr	Leu Ile Glu	His Pro Ser	Leu Lys Tyr	Ala	
118	150		155		160		165
120	ACC ACT CGT	TCC ATA GCC	AAG TAT CCT	GTG GGT ATA	GAA GTT GGT	CCT	701
121	Thr Thr Arg	Ser Ile Ala	Lys Tyr Pro	Val Gly Ile	Glu Val Gly	Pro	
122		170		175		180	
124	CAG CCT CAA	GGG GTT CTG	AGA GCT GAT	ATC TTG GAT	CAA ATG AGA	AAA	749
125	Gln Pro Gln	Gly Val Leu	Arg Ala Asp	Ile Leu Asp	Gln Met Arg	Lys	
126		185		190		195	
128	ATG ATT AAA	CAT GCT CTT	GAT TTT ATA	CAT CAT TTC	AAT GAA GGA	AAA	797
129	Met Ile Lys	His Ala Leu	Asp Phe Ile	His His Phe	Asn Glu Gly	Lys	
130		200		205		210	
134	GAA TTT CCT	CCC TGC GCC	ATT GAG GTC	TAT AAA ATT	ATA GAG AAA	GTT	845
135	Glu Phe Pro	Pro Cys Ala	Ile Glu Val	Tyr Lys Ile	Ile Glu Lys	Val	
136		215		220		225	
138	GAT TAC CCC	CGG GAT GAA	AAT GGA GAA	ATT GCT GCT	ATC ATC CAT	CCT	893

## RAW SEQUENCE LISTING

DATE: 11/19/2001

PATENT APPLICATION: US/09/965,807

TIME: 14:24:06

Input Set : N:\Crif3\RULE60\09965807.txt

Output Set: N:\CRF3\11192001\I965807.raw

```
139 Asp Tyr Pro Arg Asp Glu Asn Gly Glu Ile Ala Ala Ile Ile His Pro
140 230                235                240                245
142 AAT CTG CAG GAT CAA GAC TGG AAA CCA CTG CAT CCT GGG GAT CCC ATG      941
143 Asn Leu Gln Asp Gln Asp Trp Lys Pro Leu His Pro Gly Asp Pro Met
144                250                255                260
146 TTT TTA ACT CTT GAT GGG AAG ACG ATC CCA CTG GGC GGA GAC TGT ACC      989
147 Phe Leu Thr Leu Asp Gly Lys Thr Ile Pro Leu Gly Gly Asp Cys Thr
148                265                270                275
150 GTG TAC CCC GTG TTT GTG AAT GAG GCC GCA TAT TAC GAA AAG AAA GAA      1037
151 Val Tyr Pro Val Phe Val Asn Glu Ala Ala Tyr Tyr Glu Lys Lys Glu
152                280                285                290
154 GCT TTT GCA AAG ACA ACT AAA CTA ACG CTC AAT GCA AAA AGT ATT CGC      1085
155 Ala Phe Ala Lys Thr Thr Lys Leu Thr Leu Asn Ala Lys Ser Ile Arg
156                295                300                305
158 TGC TGT TTA CAT TAGAAATCAC TTCCAGCTTA CATCTTACAC GGTGTCTTAC      1137
159 Cys Cys Leu His
160 310
162 AAATTCTGCT AGTCTGTAAG CTCCTTAAGA GTAGGGTTGT GCCTTATTCA ACTGCATACA      1197
164 TAGCTCCTAG CACAGTGCCT TATTCGGTAG GCATCTAAGC AAATTTCTTA AATTAATTAA      1257
166 TATATCTTTA AAGATATCAT ATTTTATGTA TGTAGCTTAT TCAAAGAAGT GTTTCCTATT      1317
168 TCTATATAGT TTATTATACA TGATACTTGG GTAGCTCAAC ATTCTTAATA AACAGCCTTT      1377
170 GTATTCAGAA TATAAAATTG AAATAGATAT ATATAAAGTT AAAAAAAAAA AAAAAAAA      1435
173 (2) INFORMATION FOR SEQ ID NO: 2:
175     (i) SEQUENCE CHARACTERISTICS:
176         (A) LENGTH: 313 amino acids
177         (B) TYPE: amino acid
178         (D) TOPOLOGY: linear
181     (ix) FEATURE:
182         (A) NAME/KEY: Modified-site
183         (B) LOCATION: 83
184         (D) OTHER INFORMATION: /note= "Phosphorylation site"
186     (ix) FEATURE:
187         (A) NAME/KEY: Modified-site
188         (B) LOCATION: 105
189         (D) OTHER INFORMATION: /note= "Phosphorylation site"
191     (ix) FEATURE:
192         (A) NAME/KEY: Modified-site
193         (B) LOCATION: 108
194         (D) OTHER INFORMATION: /note= "Phosphorylation site"
198     (ix) FEATURE:
199         (A) NAME/KEY: Modified-site
200         (B) LOCATION: 146
201         (D) OTHER INFORMATION: /note= "Phosphorylation site"
203     (ix) FEATURE:
204         (A) NAME/KEY: Modified-site
205         (B) LOCATION: 264
206         (D) OTHER INFORMATION: /note= "Phosphorylation site"
208     (ix) FEATURE:
209         (A) NAME/KEY: Modified-site
```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/965,807

DATE: 11/19/2001

TIME: 14:24:06

Input Set : N:\Crif3\RULE60\09965807.txt

Output Set: N:\CRF3\11192001\I965807.raw

```

210      (B) LOCATION: 117
211      (D) OTHER INFORMATION: /note= "Potential N-glycosylation
212 site"
214      (ix) FEATURE:
215          (A) NAME/KEY: Active-site
216          (B) LOCATION: 18..24
217          (D) OTHER INFORMATION: /note= "Consensus sequence
218 predicted to be involved in catalysis"
220      (ix) FEATURE:
221          (A) NAME/KEY: Active-site
222          (B) LOCATION: 275..278
223          (D) OTHER INFORMATION: /note= "Consensus sequence
224 predicted to be involved in catalysis"
226      (ix) FEATURE:
227          (A) NAME/KEY: Active-site
228          (B) LOCATION: 283..289
229          (D) OTHER INFORMATION: /note= "Consensus sequence
230 predicted to be involved in catalysis"
233      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
235      Met Thr Ser Cys His Ile Ala Glu Glu His Ile Gln Lys Val Ala Ile
236      1          5          10          15
238      Phe Gly Gly Thr His Gly Asn Glu Leu Thr Gly Val Phe Leu Val Lys
239      20          25          30
241      His Trp Leu Glu Asn Gly Ala Glu Ile Gln Arg Thr Gly Leu Glu Val
242      35          40          45
244      Lys Pro Phe Ile Thr Asn Pro Arg Ala Val Lys Lys Cys Thr Arg Tyr
245      50          55          60
247      Ile Asp Cys Asp Leu Asn Arg Ile Phe Asp Leu Glu Asn Leu Gly Lys
248      65          70          75          80
250      Lys Met Ser Glu Asp Leu Pro Tyr Glu Val Arg Arg Ala Gln Glu Ile
251      85          90          95
253      Asn His Leu Phe Gly Pro Lys Asp Ser Glu Asp Ser Tyr Asp Ile Ile
254      100         105         110
256      Phe Asp Leu His Asn Thr Thr Ser Asn Met Gly Cys Thr Leu Ile Leu
257      115         120         125
259      Glu Asp Ser Arg Asn Asn Phe Leu Ile Gln Met Phe His Tyr Ile Lys
260      130         135         140
262      Thr Ser Leu Ala Pro Leu Pro Cys Tyr Val Tyr Leu Ile Glu His Pro
263      145         150         155         160
265      Ser Leu Lys Tyr Ala Thr Thr Arg Ser Ile Ala Lys Tyr Pro Val Gly
266      165         170         175
268      Ile Glu Val Gly Pro Gln Pro Gln Gly Val Leu Arg Ala Asp Ile Leu
269      180         185         190
271      Asp Gln Met Arg Lys Met Ile Lys His Ala Leu Asp Phe Ile His His
272      195         200         205
274      Phe Asn Glu Gly Lys Glu Phe Pro Pro Cys Ala Ile Glu Val Tyr Lys
275      210         215         220
277      Ile Ile Glu Lys Val Asp Tyr Pro Arg Asp Glu Asn Gly Glu Ile Ala
278      225         230         235         240

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/965,807

DATE: 11/19/2001

TIME: 14:24:06

Input Set : N:\Crf3\RULE60\09965807.txt

Output Set: N:\CRF3\11192001\I965807.raw

280 Ala Ile Ile His Pro Asn Leu Gln Asp Gln Asp Trp Lys Pro Leu His  
281 245 250 255  
283 Pro Gly Asp Pro Met Phe Leu Thr Leu Asp Gly Lys Thr Ile Pro Leu  
284 260 265 270  
286 Gly Gly Asp Cys Thr Val Tyr Pro Val Phe Val Asn Glu Ala Ala Tyr  
287 275 280 285  
289 Tyr Glu Lys Lys Glu Ala Phe Ala Lys Thr Thr Lys Leu Thr Leu Asn  
290 290 295 300  
292 Ala Lys Ser Ile Arg Cys Cys Leu His  
293 305 310  
295 (2) INFORMATION FOR SEQ ID NO: 3:  
297 (i) SEQUENCE CHARACTERISTICS:  
298 (A) LENGTH: 313 amino acids  
299 (B) TYPE: amino acid  
300 (D) TOPOLOGY: linear.  
303 (ix) FEATURE:  
304 (A) NAME/KEY: Region  
305 (B) LOCATION: 6  
306 (D) OTHER INFORMATION: /note= "This is isoleucine in  
307 human, valine in bovine. This is a very  
308 conservative substitution."  
310 (ix) FEATURE:  
311 (A) NAME/KEY: Region  
312 (B) LOCATION: 9  
313 (D) OTHER INFORMATION: /note= "This is glutamic acid in  
314 human, aspartic acid in bovine. This is a very  
315 conservative substitution."  
317 (ix) FEATURE:  
318 (A) NAME/KEY: Region  
319 (B) LOCATION: 10  
320 (D) OTHER INFORMATION: /note= "This is histidine in human,  
321 proline in bovine. This is a conservative  
322 substitution."  
326 (ix) FEATURE:  
327 (A) NAME/KEY: Region  
328 (B) LOCATION: 12  
329 (D) OTHER INFORMATION: /note= "This is glutamine in human,  
330 lysine in bovine. This is a very conservative  
331 substitution."  
333 (ix) FEATURE:  
334 (A) NAME/KEY: Region  
335 (B) LOCATION: 38  
336 (D) OTHER INFORMATION: /note= "This is glycine in human,  
337 serine in bovine. This is a very conservative  
338 substitution."  
340 (ix) FEATURE:  
341 (A) NAME/KEY: Region  
342 (B) LOCATION: 39  
343 (D) OTHER INFORMATION: /note= "This is alanine in human,

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/965,807

DATE: 11/19/2001

TIME: 14:24:07

Input Set : N:\Crf3\RULE60\09965807.txt

Output Set: N:\CRF3\11192001\I965807.raw

L:33 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]  
 L:34 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]  
 L:527 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
 L:533 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
 L:539 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
 L:542 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
 L:551 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
 L:563 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
 L:569 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
 L:572 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
 L:576 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
 L:579 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
 L:585 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
 L:845 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12  
 L:884 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:14  
 L:887 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:14  
 L:911 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:15  
 L:932 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:16  
 L:1162 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:25  
 L:1190 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:26  
 L:1212 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:27